''बेटी बचाओ, बेटी पढ़ाओ''

Submitted in
Partial Fulfilment for the Degree
Of
Doctor of Philosophy

(Computer Science)

Submitted to

Dr. Satish Chandra Pandey

Submitted by

Asha Singh

(JVR-I/22/10094)

[Department of Computer Science]
Jayoti Vidyapeeth Women's University, Jaipur

2023

Table of Contents

1. Introduction……3
2. Objectives……..5
3. Literature review……7
4. Research Methodology……..8
5. Data Collection……..9
6. References……11

Introduction

In any industry customer retention is one of the ever-growing concern for the company. Mostly product and services based businesses,being a dynamic and competitive industry inherently exhibits a high potential for churn,leading to the requirement of accurate churn prediction models.Customer churning is the vital issue in Customer relationship Management.

Due to the rapid advancement in electronic commerce technology, the number of competitors for each and every company are growing exponentially. Therefore, increasing the need for analysing churn rate and accordingly developing strategies for customer retention. Mostly businesses, including those in the cosmetics, clothing, electronics, applications, social media, telecommunications, video game grocery industries, worry about losing customers. Predicting when a consumer may leave a company is crucial for retaining customers and growing a business's base in every industry.

Customers have different behaviors and preferences, and reasons for cancelling their subscriptions. Therefore, churn rate prediction can help a company to take proactive measures to actively communicate with each of them to keep them on their customer list. They need to analyse and know the marketing activities that are most effective for a specific type of customers and when they are most effective.

One of the major challenges faced by churn prediction models is the high imbalance levels and high concept drift levels associated with the domain.Healthy market competition is fundamental to a well-functioning of a country’s economy. Basic economic theory demonstrates that when firms have to compete for customers, it leads to lower prices, higher quality goods and services, greater variety, and more innovation.

Shift of modern commerce and all our day-to-day activities such as grocery shopping, clothes shopping, cosmetics purchasing from offline to online modes makes the customer churn prediction necessary due to availability of too many options at a single click irrespective of the fact that online payment eases the work but also puts our life at more risk to fraud attacks such as online identity theft, banking fraud, payment card frauds, insurance frauds and money laundering.

The product competition positively influences firms' performance and increases profitability if the customer churn is taken into consideration from the very beginning and right measures are taken for customer retention accordingly.Even a small increase in customer retention can yeild a larger profit for the company than acquiring new customers.

Hence, the growing market brings in an essential need for a reliable mechanism which can efficiently predict the churn rate and keep the clients intact with the company. Data mining and machine learning helps in detecting the patterns among data attributes i.e. to predict the probability of the customer leaving or unsubscribing from the service of the company in near future.

Objectives

The purpose of research is to investigate and analyze the effectiveness of existing churn rate prediction techniques (with different intelligent methods such as machine learning, deep learning, data mining), also evaluate and comparatively analyse their performance in identifying customer churn activities. Data Mining Techniques can be used to get valuable patterns of customer behavior using customer related data.
This project also aims to propose an integrated AI-based churn rate prediction framework that combines multiple techniques to achieve higher detection rates and lower false positives and evaluate and compare the proposed framework with different techniques to know which model performs well for the real-world datasets and demonstrate its effectiveness.

Customer Churn prediction means knowing which customers are likely to leave or unsubscribe from your service. For most of the companies, this is an important prediction. This is because acquiring new customers often costs more than retaining existing ones. Once they've identified customers at risk of churn, they need to make out exactly what marketing efforts they should make with such customers to maximize their likelihood of staying.

Customer churn rate prediction is an important tool for businesses to retain customers and improve revenue. It helps to identify expected churning customer i.e. the customer who is expected to leave or unsubscribe from the service of the company in near future, so that the companies can focus on those customers instead of all customers for giving some customer specific schemes to retain those customers.

Customer churn have extremely damaged the sustainable growth of brand's market and enterprise. The amount of customer churn cases keep on increasing due to maket competition. The model aims at identifying churn rate using various data mining techniques and guard, the company's client base. This study will help to make the company's client base more secure and legitimate users more engaged with the company. Detecting the patterns of the customers who are deviating from their past history and the company take right steps to provide the customer with some attractive offers so as to rebuild the interest of the customer in the company will help the company retain its customers and secure it's client base.

Implementation of efficient churn rate prediction systems has thus become important for all sectors in order to minimize their losses by minimizing the churn rate and maximize their profits by maximizing customer retention. Many modern techniques have been suggested such as Artificial Intelligence, Machine learning, Data mining, K- Nearest Neighbour algorithm, Neural Network, Bayesian Network, Decision Tree have evolved in detecting the probable customers who are most likely to churn.

Firms like telecommunications, manufacturing and wholesale are the most affected by customer churn because of high competition resulting easy availability. As a result these industries are constantly evaluating how to stay one step ahead of their competitors. This can be quite a challenge with providing attractive offers to customers yet keep the brand value high in market. Also these risk management strategies should adapt and become more automated to cope with the increasing number of customers.

Decision Tree and Support Vector Machine can be used, if we are interested in the true churn rate and false churn rate. The Logistic Regressions might be used if we have to find the churn probability.

Literature review

A number of studies have been conducted on customer churn rate prediction. A review of the literature reveals that a variety of machine learning algorithms have been used to develop customer churn prediction models, with logistic regression, decision trees, support vector machines and artificial neural networks being the most commonly used algorithms.

Churn prediction is a critical task in various domains such as manufacturing, finance, insurance, e-commerce, healthcare, etc. With the increasing options as well as complexity and sophistication of commerce activities, traditional rule-based approaches are often insufficient to detect and prevent customer churn effectively. In recent years, the application of Artificial Intelligence (AI) techniques have shown promising results in enhancing churn prediction capabilities. This synopsis explores the use of AI for churn rate prediction and highlights its advantages in improving accuracy and efficiency.

The use of unsupervised learning techniques for churn detection have shown promising results. It has been observed that feature engineering is one of the most important step in developing effective customer churn prediction model. Since, the accuracy of the model will majorly depend on the data being fed to the model for training as well as testing. Some of the most common feature engineering techniques used in customer churn prediction include data transformation, feature selection, and feature creation.

Churn rate prediction has evolved from traditional statistical methods to incorporate advanced machine learning algorithms and varied data sources. While promising

results have been achieved, ongoing challenges ensure churn prediction remains an active research area. Despite advancements, challenges persist. Data imbalances (where the number of churned users is significantly lower than non-churned) can affect model performance. Moreover, while historical data aids prediction, rapid changes in consumer preferences can make predictions outdated.

Research Methodology

The research methodology consists of several stages. Intially, the process involves a comprehensive literature review to understand the existing methodologies and identify the latest advancements in the field, also understand the algorithms, and datasets used in churn prediction systems. Various AI techniques, including supervised learning, unsupervised learning, and pattern recognition, will be studied in detail. The research will also involve the collection and preprocessing of relevant datasets to train and evaluate the models. Evaluation metrics such as accuracy, precision, recall, and F1 score will be used to assess the performance of the models.

Steps Involved In Creating The Model :

- **Problem definition -** defining the characteristics of churn prediction systems and customer activities.
- **Data exploration -** related data is collected and explored using statistical methods to identify the underlying problems.
- **Data preparation -** data cleaning is performed which also involves finding and removing irrelevant features in the dataset to increase the prediction results and decrease model processing time. Feature selection techniques will be applied before building prediction models.

- **Modeling -** Integrated AI-based churn prediction framework will be modeled that combines multiple techniques to achieve higher prediction rates and lower false positives.
- **Evaluation -** the above model will be evaluated using different evaluation metrices such as accuracy, precision, recall, and F1 score.

  Accuracy: Accuracy is the percentage of predictions that are correct.

  Precision: Precision is percentage of positive predictions that are actually correct.

  Recall: Recall is the percentage of actual positives that are correctly predicted.

  F1 score: The F1 score is a harmonic mean of precision and recall.
- **Deployment -** making the trained model available to users so that they can make predictions on new data. It involves packaging the model, deploying it to a production environment, and integrating it with other systems.

Data Collection

The research will involve the collection of relevant datasets for training and evaluating the AI models. Real-world datasets from finance, manufacturing, e-commerce, or other suitable domains will be acquired. Data preprocessing techniques, including data cleaning, feature selection, and normalization, will be applied to ensure the quality and consistency of the data.
The data collection process will consist of the following steps:

1. Data Sources:
Multiple sources will be employed to gather relevant data on customer transactions. These sources may include Customer relationship management (CRM) systems,

Customer support databases, Website analytics, Product usage data, Financial data, Social media data. By utilizing a variety of sources, a comprehensive understanding of different types of customers and their characteristics can be obtained.

2. Data Acquisition:

The process of data acquisition will involve accessing and obtaining permission to use the required datasets. Depending on the availability and accessibility of the data, collaborations with relevant organizations and obtaining appropriate data usage agreements may be necessary. Emphasis will be placed on obtaining high-quality datasets that encompass a significant number product transactions for comparison.

3. Data Preprocessing:

Once the data has been acquired, a thorough preprocessing phase will be conducted. This step involves cleaning and transforming the data to ensure consistency and remove any inconsistencies, outliers, or missing values. Additionally, data anonymization techniques will be applied to protect the privacy and confidentiality of individuals involved in the transactions.

4. Data Variables:

Various variables will be considered during the data collection process. These variables may include transaction amounts, timestamps, geographical information, customer demographics, behavioral patterns, and other relevant features. The inclusion of a wide range of variables will enable a comprehensive analysis of fraud patterns and aid in the development of effective churn prediction model.

5. Data Sample:

To ensure statistical validity and representativeness of the findings, a carefully selected sample from the collected data will be used for analysis. The sample will be representative of the overall population and will account for different types of

customer trends such as transaction volumes and time periods. The sample size will be determined based on statistical considerations and the specific research objectives. By employing a rigorous data collection process encompassing diverse sources and robust preprocessing techniques, this study aims to gather a comprehensive dataset for the analysis methods and the development of effective churn prediction model.

References

1. Sudharsan, R . (2022). Big Data Analytics in Telecommunication for Customer Churn Prediction.
2. Kapoor , V. (2022). Mining Mobile Data for Telecom Churn Management.
3. Isabella Amali , A. (2021). an extensive performance analysis on framework of customer churn prediction through stochastic gradient descent based deep learning model.
4. KAMALRAJ , N. (2017, May 12). Design And Development Of Decision Tree Based Techniques For Churn Prediction In Service Industry.